INTSUM in terms of the underlying structural features around the bonds in question that seem to "drive" the fragmentations. For example, INTSUM will notice significant fragmentation of the two different bonds alpha to the carbonyl group in aliphatic ketones. It is left to RULEGEN to discover that these are both instances of the same fundamental alpha-cleavage process that can be predicted any time a bond is alpha to a carbonyl group.

The RULEMOD program modifies and condenses the set of rules produced by INTSUM and RULEGEN together. It looks at the negative evidence associated with each candidate rule in order to select the best ones, then merges rules that seem to explain the same breaks (if possible). The program was substantially improved in several ways, as described in the next section.

2.3.1 INTSUM Improvements

Transfers of arbitrary neutral species can now be specified as part of the mass spectrometry processes, instead of transfers of hydrogen atoms alone. This capability increases the utility of the program in at least two ways: first, it allows a chemist to control the program better -- to produce the kinds of results that are more chemically meaningful -- and second, it allows the program to explore more complex processes within its space and time limitations. For example, carbon monoxide and water were listed as plausible neutral molecules to transfer in or out of fragments for the triketoandrostanes. Thus, the processes are listed with and without these transfers, just as chemists prefer, instead of showing loss of CO as a set of two breaks around the keto group, or loss of H_2O as loss of oxygen (breaking the C=O bond) accompanied by loss of two hydrogens. What is more, the program can now produce these results without violating its chemical heuristics of (a) not breaking adjacent bonds, and (b) not breaking double bonds. This economy also pays off in increasing the complexity of the processes that can be considered. Because loss of CO, for example, is a result of a transfer instead of the result of breaking two bonds, the number of bonds broken in accompanying processes can be increased by two.

Another INTSUM improvement was to increase the options for initial data filtering. Thresholding is too simple for many problems, so we now provide an option to cluster peaks and select the n largest peaks from each cluster.

The format of the input data is also now less strict than before. We have written programs to read spectra in Aldermaston format. And we have merged CONGEN's Editstruc package into the INTSUM setup routines to allow a chemist to associate structures with spectra interactivity. This greatly decreases the chances of error in setting up the input data.

Several modifications were also made to the program to increase its efficiency, e.g., processing all intensities as integers (between 0 and 1000).

2.3.2 RULEGEN Improvements

The evaluation of prospective rules in RULEGEN guides the entire rule generation procedure. To tune this procedure, we modified the evaluation function in several ways and compared the resulting sets of rules. We were looking for an objective way of telling the program to keep rules general, but "not too general". The current evaluation function is substantially improved as a result.

Because the RULEGEN program searches such a large space of partial and complete rules, it requires large amounts of computer time (sometimes more than 60 cpu minutes). Thus, we have investigated several improvements for efficiency alone. In addition, we have made the program easier to set up and run in batch mode to reduce the chemist's personal time investment. And we have made the program easily restarted from any intermediate point -- to protect the chemist from machine failures.

2.3.3 RULEMOD Improvements

At the time of the last annual report RULEMOD was a new program still in its experimental stages. Since then we have added new subprograms and integrated the program with other programs to make it a useful and necessary part of Meta-DENDRAL.

Two new subprograms greatly improve RULEMOD's performance. (1) A program to add specifications to rules was completed. It looks for plausible ways of making a rule more specific in order to decrease the number of counterexamples to the rule. (2) A complementary program to make rules more general was also completed. The program tries to find ways to reduce the number of descriptors on nodes of subgraphs in order to increase the breadth of applicability of rules. Its major constraint is that it cannot make any change that would increase the number of counterexamples. Both of these subprograms make the final rules much closer to rules that chemists approve of.

The subprogram that merges rules was also improved. The program tries to merge pairs of rules into a more general form for economy and clarity of rules. Its major constraint is that no explanations are lost, *i.e.*, all the data points explained by the initial pair of rules will still be explained after merging. Formerly we insisted that the more general form must cover all the same data points as the initial rules, but this was found to be too narrow a constraint. By giving the program a more global view of the entire set of rules, we can let the more general, merged form explain fewer data points as its component rules as long as other rules explain the remainder.

PART 3: APPLICATIONS TO BIOMEDICAL STRUCTURE ELUCIDATION PROBLEMS

3.1 Introduction

In our grant proposal we discussed the application of the instrumentation and computer programs described above to the study of molecular structure problems in a variety of biomedical applications areas. This is our primary research area, and we discussed specific classes of problems and compounds for investigation. We also made it quite clear that our facilities would be made available to wider community of collaborators/users as our resources permitted. Both categories of application, i.e., within our own group, and with an outside group, are described in some detail below.

Our last annual report described several steps taken to encourage a broad community of researchers to use our facilities. For example, we sent a questionnaire to members of the American Society for Mass Spectrometry, Committee III on Computer Applications, and a follow-up letter to persons indicating a desire to know more about access to our programs. The same note has been sent to several other persons whom we know from personal contacts might be interested. Because of the nature of their investigations, many of these people receive NIH support. Several of our publications (e.g., [45]-[49]) mention the availability of our programs. In addition, through individual contacts and formal presentations at conferences we have been encouraging outside use of the programs.

The availability of SUMEX as a mechanism for resource sharing has made it possible for us to extend access to our programs to a number of people. Without SUMEX, this access would be impossible, and most of our programs (those which are not easily exportable) could be used only by ourselves.

3.2 Applications by Professor Djerassi's Research Group

Our existing grants, outlined below, mesh well with our instrumentation and program development under the present award. Under NIH Grant GM06840 we have been studying natural products from marine sources with major emphasis on terpenoids and sterols. For this work we have been dependent on the use of our 711 instrument for high resolution mass spectrometry which we require for the identification of all new compounds, many of which are present in only very small quantities. We are particularly anxious to have access to GC coupled with a high resolution mass spectrometer because we hope to be able to screen large numbers of marine animals for their sterol content using this technique.

We are currently engaged in intensive efforts in analysis of mixtures of marine sterols involving our computer-based procedures. The program for the development of the computer operated and assisted system of marine sterol structure analysis has been planned to proceed in three stages:

1) Analysis of all literature published concerning marine sterols so that a complete listing of known sterol structures and organisms studied could be compiled.

2) Collection, evaluation, digitization and computer file construction for the mass spectra of all known marine sterols, followed by the institution of a computer operated file search sequence for direct analysis of marine sterol GC-MS data.

3) The application of the INTSUM, RULEGEN, and RULEMOD programs to the computer file of marine sterol spectra so that a series of fragmentation rules can be extracted for use in the generation of possible structures from mass spectral data for new marine sterols, that is, sterols whose mass spectra cannot be matched with any spectra contained in the computer search file.

3.3 Applications of Programs by External Scientists

The DENDRAL project, still one of the major users of the SUMEX-AIM computer facility, has formed a small community of regular, remote users. This "exodendral" community has continued to provide valuable contributions to program development, although the growth of this community has had to be slowed in response to increasing demands by other projects upon the SUMEX-AIM facility. As an example, for the months of September 1975 to February 1976, the number of CPU hours used by exodendral persons amounted to at least 8 percent of the CPU hours used by the DENDRAL project. There are currently four remote chemist-users whose groups´ regularly use CONGEN in their day to day work. Additionally, there are several remote users who use their accounts on an occasional basis, or who access SUMEX-AIM via the GUEST mechanism.

The SUMEX-AIM facility has grown markedly in number of projects over the past year. Due to this increase in system loading; the DENDRAL project, which had previously been able to offer trial usage of its programs to almost any chemist who expressed a need to use the programs, has found itself in the unfortunate position of of having to carefully screen potential collaborators. Those chemists who have been granted access, have been requested to restrict their usage to off-prime time hours. CONGEN, the DENDRAL program which receives most of this usage, has evolved in a manner designed to try to remedy the system loading problem which can be created by the enthusiasm of it´s chemist-users. Since a typical, long GENERATE, PRUNE or IMBED within CONGEN can be very time consuming, as well as a voracious consumer of CPU cycles, a provision to permit a user to easily take advantage of SUMEX-AIM´s off-hour batch processing has been implemented. A CONGEN user can now interactively set up his problem, and when ready to commence with a time consuming procedure, can, from within CONGEN, request automatic submission to BATCH, to be run late at night. The CONGEN users´ also benefit from this ability, in that they no longer must leave a terminal tied up during the sometimes hour-long compute times. This development then, can be viewed as responding to CONGEN users´ needs as well as being an effort by the DENDRAL project to be conscientious in its resource-sharing responsibilities.

Following is a brief summary of the major users of CONGEN over the past year, as well as notes on chemists who contacted us about trial usage of the programs.

Dr. Clair Cheer, Professor of Chemistry, University of Rhode Island, Kingston, Rhode Island. Dr. Cheer is on sabbatical leave from the University of Rhode Island to the Stanford University Chemistry Department. He has, in recent work with Professor Djerassi´s group, demonstrated the utility of CONGEN in the identification of (+)-Palustrol, a tricyclic sesquiterpene alcohol from the marine Xeniid Cespitularia virdis (Cheer, Djerassi et. al., Tetrahedron, in press). Dr. Cheer plans to continue his work with CONGEN once he returns to Rhode Island in December.

Dr. Jon Clardy, Professor of Chemistry, Iowa State University. Dr. Clardy read of CONGEN in an article appearing in the Journal of the American Chemical Society and contacted Professor Djerassi concerning the possibility of using the program from Iowa. He was offered GUEST access during the winter of 1975, but has not yet had an opportunity to evaluate the potentials of the program.

Dr. Douglas Dorman, Eli Lily Corp., Indianapolis, Indiana. Dr. Dorman's research involves the identification and characterization of drug related compounds by chemical and spectrographic methods. Using primarily the NMR and C13 NMR spectra of these various compounds, Dr. Dorman has found CONGEN to be a time-saving adjunct to his structure elucidation work.

Dr. H.M. Fales, National Heart and Lung Institute, Bethesda, Maryland. Dr. Fales, along with Doctors Sanford Markey and Peter Roller had a joint account set up for them in April of 1975. Most of the use of this account came during late summer at which time Dr. Fales experimented with the use of CONGEN for assistance in the elucidation of the structure of a novel quinolinone, known to be tumorogenic. Although the crystal structure had been solved at the time of his usage of CONGEN, Dr. Fales felt that the program produced an abundance of useful ideas. The main problem initially faced by Dr. Fales in using CONGEN was in getting a feel for problem size and the effects of various constraint types.

Professor Kenneth Gash, California State College at Dominguez Hills. Professor Gash is a professor of chemistry who is on temporary leave to Small College, the research branch of Dominguez Hills. Dr. Gash did some of the original work, in 1965, with Professor Morton Munk, on the structure elucidation program developed at Arizona State University. Dr. Gash has been reviewing some of the problems originally done with Munk's program and has been studying input, output and constraint capabilities found in CONGEN. He has generally concluded that CONGEN provides an excellent tool for the chemist to use in structure elucidation problems subject to the constraint of slow system response time.

Mr. Neil A. B. Gray, King's College, Cambridge, England. Mr. Gray, following a three week visit to the Stanford chemistry department, requested copies of all the current DENDRAL programs to be sent to him in England. He is a chemist who has been working in areas related to developments in various of the DENDRAL programs, and hopes to be able to benefit from work already done at Stanford. His current interest in intelligent constraint application during structure elucidation merges well with one of the directions in which CONGEN is tending to develop. Unfortunately, Mr. Gray does not have access to an ARPANET or TYMNET node to access SUMEX-AIM directly. Therefore, all collaboration has had to be carried on by mail.

Dr. Jerrold Karliner, Ciba Geigy Corporation, Ardsley, New York. Dr. Karliner and his research group at Ciba-Geigy have become regular users of CONGEN in their day-to-day operation of a research laboratory. Dr. Karliner is a completely self-taught user of CONGEN, and has served to encourage others to request permission to use this program.

Dr. Milton Levenberg, Abbott Laboratories, Chicago, Illinois. Dr. Levenberg has been an occasional user of CONGEN as an adjunct to his work as head of a mass-spectrometry laboratory. Primary usage has been to provide assurance that the proposal of a structure for a compound on the basis of chemical and spectroscopic evidence has not overlooked other plausible possibilities.

Dr. Gino Marco, Ciba Geigy Corporation, Greensboro, North Carolina. Dr. Marco heard about CONGEN during a company seminar presented by Dr. Karliner. After a brief trial use via the GUEST mechanism, Dr. Marco requested an account for use by his group of metabolic and organic chemists. Dr. Marco's research group studies unknown insect metabolites by micro-IR and micro-NMR methods, and attempts structure elucidation based on these forms of spectroscopic analysis. Testing the utility of the program before implementing it for day to day use, Dr. Marco discovered that CONGEN could greatly narrow the alternatives of complex metabolic conjugates which had to be considered in a typical elucidation problem. They have established a leased line to the nearest TYMNET node, and expect increased CONGEN usage in the future.

Professor G. Minole, Italy. Professor Minole has been active in elucidation of structures of marine natural products, an area of interest which overlaps with our own. We have provided, by written communication due to absence of network access, sets of structural alternatives in current problems being studied by Professor Minole. We have used some of the mass spectrometric prediction functions of our DENDRAL programs to determine which structures in a set of possibilities could yield the observed mass spectral data.

Professor Nogi Nakanishi, Department of Chemistry, Columbia University. Professor Nakanishi is one of the most active and productive persons engaged in structure elucidation activities. He has developed an active interest in CONGEN and is collaborating with us on several novel problems. One of these problems has involved the structure of the active component of defense secretions of an insect (termite). Other defense secretion components are under investigation as we explore structural alternatives based on current data.

Dr. David Pensak, DuPont de Nemours and Company, Wilmington, Delaware. Indirectly requested information about CONGEN through a letter written by his immediate superior to Professor Lederberg. Dr. Pensak has been offered GUEST access, and has just begun a potential collaboration with a DENDRAL group which is studying model builders and their production of reliable geometries for certain types of molecules.

Professor Manfred Wolff, University of California at San Francisco. Dr. Wolff is chairman of the Department of Pharmacological Chemistry, and inquired as to the possibilities of accessing SUMEX-AIM and appropriate programs for a faculty which is interested in many aspects of drug design and drug action, ranging from physical chemistry to purely biological studies. He has been encouraged to use GUEST access to explore CONGEN, although he has taken no action up to the present time.

We have cases where requests for GUEST access had to be denied due to system loading considerations. We made these decisions according to the extent to which the requested use would fit within the research guidelines of SUMEX/AIM and our own stated criteria from the 1973 proposal to NIH. In one case, for instance, the use was for an individual's report on potential educational uses of CONGEN.

FUNDING STATUS

The DENDRAL project is in its sixth year of NIH funding through the BRB (Grant RR-00612). For the period 8/1/75 - 7/31/76 the total (direct costs) amount awarded was $240,967. After nine months of the seventh year the project cease to be supported by the current grant: a competing renewal application will be submitted June 1, 1976. For the nine months period 8/1/76 - 4/30/77 the total (direct costs) amount awarded is $210,778.

INTERACTIONS WITH THE SUMEX-AIM RESOURCE

The research summary above described several ways in which we see the DENDRAL programs helping biomedical scientists. See Part 3 for a list of persons with whom we have actively collaborated. One of the major goals of the research is to extend the usefulness of the programs for just such persons.

The SUMEX-AIM community is an exciting and productive collection of projects and individuals who contribute in many ways to the progress of all projects in the community. Our programming in INTERLISP, SAIL and FORTRAN, for example, is speeded considerably by the ready availability of expert programmers from many projects. We have shared ideas about intelligent interfaces between programs and users with members of the MYCIN, X-Ray Crystallography and MOLGEN projects. Perhaps the most used and most useful means of communication is the SNDMSG program on SUMEX. It is much more efficient than campus mail and much less intrusive, as well as more efficient for multiple messages, than the telephone. We are cooperating with the SUMEX staff on the Bulletin Board facility, which will be another efficient means of communicating, especially when the sender of a message is not certain who the receivers should be. (It will allow potential receivers to say what they are interested in and notify them of relevant bulletins, without the sender making an explicit distribution list).

The SUMEX-AIM staff is the most professional computer facility staff we have worked with over the ten year life of the DENDRAL project. The very low amount of unscheduled downtime is a direct indication of their professional attitude and abilities. Less measurably, the helpfulness of the staff also translates directly into increased productivity for

DENDRAL. There have been numerous instances of the SUMEX staff answering our questions immediately and fixing errors in system programs for us as quickly as we could expect.

As the system becomes more heavily loaded, we notice longer and longer delays in computer response time. This is the one major criticism voiced by DENDRAL project members. Many of these persons have changed their work habits to conform to the lighter loading between midnight and 5:00 because they cannot get any significant computing done during the day.

SUMMARY OF PUBLICATIONS

(1) J. Lederberg, "DENDRAL-64 - A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs", (technical reports to NASA, also available from the author and summarized in (12)). (1a) Part I. Notational algorithm for tree structures (1964) CR.57029 (1b) Part II. Topology of cyclic graphs (1965) CR.68898 (1c) Part III. Complete chemical graphs; embedding rings in trees (1969)

(2) J. Lederberg, "Computation of Molecular Formulas for Mass Spectrometry", Holden-Day, Inc. (1964).

(3) J. Lederberg, "Topological Mapping of Organic Molecules", Proc. Nat. Acad. Sci., 53:1, January 1965, pp. 134-139.

(4) J. Lederberg, "Systematics of organic molecules, graph topology and Hamilton circuits. A general outline of the DENDRAL system." NASA CR-48899 (1965)

(5) J. Lederberg, "Hamilton Circuits of Convex Trivalent Polyhedra (up to 18 vertices), Am. Math. Monthly, May 1967.

(6) G. L. Sutherland, "DENDRAL - A Computer Program for Generating and Filtering Chemical Structures", Stanford Artificial Intelligence Project Memo No. 49, February 1967.

(7) J. Lederberg and E. A. Feigenbaum, "Mechanization of Inductive Inference in Organic Chemistry", in B. Kleinmuntz (ed) Formal Representations for Human Judgment, (Wiley, 1968) (also Stanford Artificial Intelligence Project Memo No. 54, August 1967).

(8) J. Lederberg, "Online computation of molecular formulas from mass number." NASA CR-94977 (1968)

(9) E. A. Feigenbaum and B. G. Buchanan, "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry", in Proceedings, Hawaii International Conference on System Sciences, B. K. Kinariwala and F. F. Kuo (eds), University of Hawaii Press, 1968.

(10) B. G. Buchanan, G. L. Sutherland, and E. A. Feigenbaum, "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic

Chemistry". In Machine Intelligence 4 (B. Meltzer and D. Michie, eds) Edinburgh University Press (1969), (also Stanford Artificial Intelligence Project Memo No. 62, July 1968).

(11) E. A. Feigenbaum, "Artificial Intelligence: Themes in the Second Decade". In Final Supplement to Proceedings of the IFIP68 International Congress, Edinburgh, August 1968 (also Stanford Artificial Intelligence Project Memo No. 67, August 1968).

(12) J. Lederberg, "Topology of Molecules", in The Mathematical Sciences - A Collection of Essays, (ed.) Committee on Support of Research in the Mathematical Sciences (COSRIMS), National Academy of Sciences - National Research Council, M.I.T. Press, (1969), pp. 37-51.

(13) G. Sutherland, "Heuristic DENDRAL: A Family of LISP Programs", to appear in D. Bobrow (ed), LISP Applications (also Stanford Artificial Intelligence Project Memo No. 80, March 1969).

(14) J. Lederberg, G. L. Sutherland, B. G. Buchanan, E. A. Feigenbaum, A. V. Robertson, A. M. Duffield, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference I. The Number of Possible Organic Compounds: Acyclic Structures Containing C, H, O and N". Journal of the American Chemical Society, 91:11 (May 21, 1969).

(15) A. M. Duffield, A. V. Robertson, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, "Application of Artificial Intelligence for Chemical Inference II. Interpretation of Low Resolution Mass Spectra of Ketones". Journal of the American Chemical Society, 91:11 (May 21, 1969).

(16) B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, "Toward an Understanding of Information Processes of Scientific Inference in the Context of Organic Chemistry", in Machine Intelligence 5, (B. Meltzer and D. Michie, eds) Edinburgh University Press (1970), (also Stanford Artificial Intelligence Project Memo No. 99, September 1969).

(17) J. Lederberg, G. L. Sutherland, B. G. Buchanan, and E. A. Feigenbaum, "A Heuristic Program for Solving a Scientific Inference Problem: Summary of Motivation and Implementation", Stanford Artificial Intelligence Project Memo No. 104, November 1969.

(18) C. W. Churchman and B. G. Buchanan, "On the Design of Inductive Systems: Some Philosophical Problems". British Journal for the Philosophy of Science, 20 (1969), pp. 311-323.

(19) G. Schroll, A. M. Duffield, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, "Application of Artificial Intelligence for Chemical Inference III. Aliphatic Ethers Diagnosed by Their Low Resolution Mass Spectra and NMR Data". Journal of the American Chemical Society, 91:26 (December 17, 1969).

(20) A. Buchs, A. M. Duffield, G. Schroll, C. Djerassi, A. B. Delfino, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J.

Lederberg, "Applications of Artificial Intelligence For Chemical Inference. IV. Saturated Amines Diagnosed by Their Low Resolution Mass Spectra and Nuclear Magnetic Resonance Spectra", Journal of the American Chemical Society, 92, 6831 (1970).

(21) Y.M. Sheikh, A. Buchs, A.B. Delfino, G. Schroll, A.M. Duffield, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference V. An Approach to the Computer Generation of Cyclic Structures. Differentiation Between All the Possible Isomeric Ketones of Composition C6H10O", Organic Mass Spectrometry, 4, 493 (1970).

(22) A. Buchs, A.B. Delfino, A.M. Duffield, C. Djerassi, B.G. Buchanan, E.A. Feigenbaum and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference VI. Approach to a General Method of Interpreting Low Resolution Mass Spectra with a Computer", Chem. Acta Helvetica, 53, 1394 (1970).

(23) E.A. Feigenbaum, B.G. Buchanan, and J. Lederberg, "On Generality and Problem Solving: A Case Study Using the DENDRAL Program". In Machine Intelligence 6 (B. Meltzer and D. Michie, eds.) Edinburgh University Press (1971). (Also Stanford Artificial Intelligence Project Memo No. 131.)

(24) A. Buchs, A.B. Delfino, C. Djerassi, A.M. Duffield, B.G. Buchanan, E.A. Feigenbaum, J. Lederberg, G. Schroll, and G.L. Sutherland, "The Application of Artificial Intelligence in the Interpretation of Low-Resolution Mass Spectra", Advances in Mass Spectrometry, 5 (1971), 314.

(25) B.G. Buchanan and J. Lederberg, "The Heuristic DENDRAL Program for Explaining Empirical Data". In proceedings of the IFIP Congress 71, Ljubljana, Yugoslavia (1971). (Also Stanford Artificial Intelligence Project Memo No. 141.)

(26) B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, "A Heuristic Programming Study of Theory Formation in Science." In proceedings of the Second International Joint Conference on Artificial Intelligence, Imperial College, London (September, 1971). (Also Stanford Artificial Intelligence Project Memo No. 145.)

(27) Buchanan, B. G., Duffield, A.M., Robertson, A.V., "An Application of Artificial Intelligence to the Interpretation of Mass Spectra", Mass Spectrometry Techniques and Appliances, Edited by George W. A. Milne, John Wiley & Sons, Inc., 1971, p. 121-77.

(28) D.H. Smith, B.G. Buchanan, R.S. Engelmore, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference VIII. An approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids", Journal of the American Chemical Society, 94, 5962-5971 (1972).

(29) B.G. Buchanan, E.A. Feigenbaum, and N.S. Sridharan, "Heuristic Theory Formation: Data Interpretation and Rule Formation". In Machine Intelligence 7, Edinburgh University Press (1972).

(30) Lederberg, J., "Rapid Calculation of Molecular Formulas from Mass Values". Jnl. of Chemical Education, 49, 613 (1972).

(31) Brown, H., Masinter L., Hjelmeland, L., "Constructive Graph Labeling Using Double Cosets". Discrete Mathematics, 7 (1974), 1-30. (Also Computer Science Memo 318, 1972).

(32) B. G. Buchanan, Review of Hubert Dreyfus´ "What Computers Can´t Do: A Critique of Artificial Reason", Computing Reviews (January, 1973). (Also Stanford Artificial Intelligence Project Memo No. 181)

(33) D. H. Smith, B. G. Buchanan, R. S. Engelmore, H. Adlercreutz and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens". Journal of the American Chemical Society 95, 6078 (1973).

(34) D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, C. Djerassi and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference X. Intsum. A Data Interpretation Program as Applied to the Collected Mass Spectra of Estrogenic Steroids". Tetrahedron, 29, 3117 (1973).

(35) B. G. Buchanan and N. S. Sridharan, "Rule Formation on Non-Homogeneous Classes of Objects". In proceedings of the Third International Joint Conference on Artificial Intelligence (Stanford, California, August, 1973). (Also Stanford Artificial Intelligence Project Memo No. 215.)

(36) D. Michie and B.G. Buchanan, "Current Status of the Heuristic DENDRAL Program for Applying Artificial Intelligence to the Interpretation of Mass Spectra". August, 1973. To appear in Computers for Spectroscopy (ed. R.A.G. Carrington) London: Adam Hilger. Also: University of Edinburgh, School of Artificial Intelligence, Experimental Programming Report No. 32 (1973).

(37) H. Brown and L. Masinter, "An Algorithm for the Construction of the Graphs of Organic Molecules", Discrete Mathematics, 8(1974), 227. (Also Stanford Computer Science Dept. Memo STAN-CS-73-361, May, 1973)

(38) D.H. Smith, L.M. Masinter and N.S. Sridharan, "Heuristic DENDRAL: Analysis of Molecular Structure," Proceedings of the NATO/CNNA Advanced Study Institute on Computer Representation and Manipulation of Chemical Information (W. T. Wipke, S. Heller, R. Feldmann and E. Hyde, eds.) John Wiley and Sons, Inc., 1974.

(39) R. Carhart and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference XI: The Analysis of C13 NMR Data for Structure Elucidation of Acyclic Amines", J. Chem. Soc. (Perkin II), 1753 (1973).

(40) L. Masinter, N.S. Sridharan, R. Carhart and D.H. Smith, "Application of Artificial Intelligence for Chemical Inference XII: Exhaustive Generation of Cyclic and Acyclic Isomers". Journal of the American Chemical Society, 96 (1974), 7702. (Also Stanford Artificial Intelligence Project Memo No. 216.)

(41) L. Masinter, N.S. Sridharan, R. Carhart and D.H. Smith, "Applications of Artificial Intelligence for Chemical Inference. XIII. Labeling of Objects having Symmetry". Journal of the American Chemical Society, 96 (1974), 7714.

(42) N.S. Sridharan, Computer Generation of Vertex Graphs, Stanford CS Memo STAN-CS-73-381, July, 1973.

(43) N.S. Sridharan, et.al., A Heuristic Program to Discover Syntheses for Complex Organic Molecules, Stanford CS Memo STAN-CS-73-370, June, 1973. (Also Stanford Artificial Intelligence Project Memo No. 205.)

(44) N.S. Sridharan, Search Strategies for the Task of Organic Chemical Synthesis, Stanford CS Memo STAN-CS-73-391, October, 1973. (Also Stanford Artificial Intelligence Project Memo No. 217.)

(45) R. G. Dromey, B. G. Buchanan, J. Lederberg and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference. XIV. A General Method for Predicting Molecular Ions in Mass Spectra". Journal of Organic Chemistry, 40 (1975), 770.

(46) D. H. Smith, "Applications of Artificial Intelligence for Chemical Inference. XV. Constructive Graph Labelling Applied to Chemical Problems. Chlorinated Hydrocarbons". Analytical Chemistry, in press (to appear May or June, 1975).

(47) R. E. Carhart, D. H. Smith, H. Brown and N. S.Sridharan, "Applications of Artificial Intelligence for Chemical Inference. XVI. Computer Generation of Vertex Graphs and Ring Systems". Journal of Chemical Information and Computer Science (formerly Journal of Chemical Documentation), in press (to appear in May, 1975).

(48) R. E. Carhart, D. H. Smith, H. Brown and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference. XVII. An Approach to Computer-Assisted Elucidation of Molecular Structure". Journal of the American Chemical Society, submitted for publication.

(49) B. G. Buchanan, "Scientific Theory Formation by Computer." To appear in Proceedings of NATO Advanced Study Institute on Computer Oriented Learning Processes, 1974, Bonas, France.

(50) E. A. Feigenbaum, "Computer Applications: Introductory Remarks," in Proceedings of Federation of American Societies for Experimental Biology, Vol. 33, No. 12 (Dec., 1974) 2331-2332.

(51) S. Hammerum and C. Djerassi, "Mass Spectrometry in Structural and Stereochemical Problems - CCXLIV; The Influence of Substituents and

Stereochemistry on the Mass Spectral Fragmentation of Progesterone." Tetrahedron (accepted for publication), 1975.

(52) S. Hammerum and C. Djerassi, "Mass Spectrometry in Structural and Stereochemical Problems CCXLV. The Electron Impact Induced Fragmentation Reactions of 17-Oxygenated Progesterones." Steroids (submitted for publication).

(53) H. Brown, "Molecular Structure Elucidation III." Submitted for publication to SIAM Journal on Computing.

(54) R. Davis and J. King, "Overview of Production Systems" To appear in Machine Representation of Knowledge, Proceedings of the NATO ASI Conference, July, 1975.

(55) B. G. Buchanan, "Applications of Artificial Intelligence to Scientific Reasoning." In Proceedings of Second USA-Japan Computer Conference, August, 1975.

(56) R. E. Carhart, S. M. Johnson, D. H. Smith, B. G. Buchanan, R. G. Dromey, J. Lederberg, "Networking and a Collaborative Research Community: A Case Study Using the DENDRAL Program," to appear in Computing Networking in Chemistry, Peter Lykos, ed., American Chemical Society Symposium Series, No. 19, 1975.

(57) D. H. Smith (Paper XVIII) "The Scope of Structural Isomerism". Journal of Chemical Information and Computer Sciences, 15, 203 (1975).

(58) B. G. Buchanan, D. H. Smith, W. C. White, R. Gritter, E. A. Feigenbaum, J. Lederberg and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference. XXII. Automatic Rule Formation in Mass Spectrometry by Means of the Meta-DENDRAL Program." Submitted to Journal of the American Chemical Society.

(59) E. H. Shortliffe, R. Davis, S. G. Axline, B. G. Buchanan, C. C. Green and S. N. Cohen, "Computer-Based Consultations in Clinical Therapeutics: Explanation and Rule Acquisition Capabilities of the MYCIN System." Computers and Biomedical Research 8, 303-320 (1975).

(60) R. Davis, B. Buchanan and E. Shortliffe, "Production Rules as a Representation for a Knowledge-Based Consultation Program", accepted for publication by Artificial Intelligence. (Also Stanford Artificial Intelligence Project Memo No. AIM-266.)

IV.A.1.b MYCIN PROJECT

Computer Based Consultation in Clinical Therapeutics

Prof. S. Cohen, M.D. (Pharmacology) and
Dr. B. Buchanan (Computer Science)

(Grant HEW HSO-1544-02, 3 years, $163,965 this year)

Introduction

This report offers a review of the progress made by the MYCIN project over the past year. To provide some background, we start by describing the system's basic task, and document its significance. This is followed by a description of the way knowledge is represented and used in the system, and a brief discussion of the advantages of the representation we have chosen. The progress report follows this, detailing the accomplishments of the past twelve months, and spells out the plans for the coming year.

Background

The ultimate aim of the MYCIN project has been to develop a computer-based system to which physicians will refer for antimicrobial therapy advice. One primary consideration for the system has been its level of performance. In order to provide a tool which would actually be useful, and be used in the clinical setting, we have to provide a system which displays a high level of competence in its field. Clinicians must have confidence in a program's ability before they will be willing to use it. A second consideration has been the ability of the system to explain its reasoning. Since clinicians are not likely to accept such a system unless they can understand why the recommended therapy has been selected, the system has to do more than give dogmatic advice. It is also important to let it explain its recommendations when queried, and to do so in terms that suggest to the physician that the program approaches problems in much the same way that he does. This permits the user to validate the program's reasoning, and to reject the advice if he feels that a crucial step in the decision process cannot be justified. It also gives the program an inherent instructional capability, allowing the physician to learn from each consultation session. Third, we feel it is desirable that an expert in infectious disease therapy who notes omissions or errors in the program's reasoning should be able to augment or correct the knowledge base so that future consultations will not repeat the same mistakes. The system should therefore have some capability for acquiring knowledge via interaction with experts in the field.

Progress towards these goals has been made in development of the MYCIN system, composed of three interrelated modules. The Consultation System uses MYCIN's knowledge base along with patient data entered by the physician to generate therapeutic advice. The Explanation System has the ability to generate a thorough documentation of the motivation for questions the system asks or of the rationale for conclusions it reaches. Finally, experts may use the Rule Acquisition System to update MYCIN's knowledge base. Together, these three modules give the system a wide range of capabilities for dealing with the problem of advising on diagnosis and therapy selection for infectious disease.

Significance of the problem

The task of therapy selection for infectious disease was chosen because of the demonstrated need for high quality advice in this area. There have been numerous studies detailing the misuse of antibiotics and its resultant cost. One study (reference [2]) indicates that in a recent year, one of every four people in this country were given penicillin, and nearly 90% of these prescriptions were unnecessary. A major evaluation of the antibiotic prescribing habits of a wide range of specialists was reported within the last two months in reference [3]. It indicates that the overall score was only 68% correct, and suggests possible underlying causes. While there are a number of sociological factors which are also significant (e.g. patient pressure for treatment even when none is indicated), the study suggests that causes for the low score range from the fact that physicians may be unfamiliar with generic names for antibiotics, to a lack of knowledge of basic bacteriology, to the fact that they appear to use antibiotics as a substitute for clinical judgment. Problems such as these indicate the need for more (and more accessible) consultants to physicians selecting antimicrobial drugs.

General Approach

To give a general feeling for the way MYCIN works, we present here a brief description of the way knowledge is represented in the program, and indicate how it is used. We also suggest some of the advantages which result from embodying knowledge in the format we have chosen.

All knowledge used by MYCIN during a consultation session is contained in decision rules that have been coded and stored in the machine. The MYCIN Project members have identified approximately 400 such rules during discussions of representative case histories. Each rule consists of a set of preconditions (called a PREMISE) which, if true, permits a conclusion to be made or an action to be taken, according to the ACTION part of the rule. Figure 1 below shows one such rule.

If

1) the stain of the organism is gramnegative, and
2) the morphology of the organism is rod, and
3) the aerobicity of the organism is aerobic,

Then

there is suggestive evidence (.6) that the identity of the organism is bacteroides.

RULE124

Figure 1

The system uses its collection of rules to make its conclusions. If, for instance, it is attempting to determine the identity of an organism which is causing an infection, it retrieves the entire list of rules which, like the one above, conclude about identity. It then attempts to determine the truth of the premise of the first rule on the list by evaluating in turn each of the clauses of its premise. Thus, for the rule above, the first thing to find out is the gramstain. If this is already available in the data base, the program retrieves it from there. If not, gramstain becomes the new goal, we retrieve all rules which conclude about it, and try to use each of them to obtain the value of the gramstain. If, after trying all the rules on the list, the answer still has not been discovered, the program asks the user. The rules thus "unwind" to produce a succession of goals, and it is the attempt to achieve each goal that drives the consultation. Figure 2 below gives a graphical view of this process. (A more complete description of the program's operation can be found in reference [1]).

```
                    |               |
                    |   identity    |
                    |               |
                       /   |   \
                      /    |    \
                     /     |     \
                    /      |      \
               RULE124    other rules......
               / |   \
              /  |    \
             /   |     \
      gramstain  morphology  aerobicity
```

Figure 2

Many of the system's important capabilities are made possible by way knowledge is represented in rules like the one in Figure 1. Such rules offer modular "chunks" of knowledge about the domain, represented in a form that is comprehensible to the clinician. For instance, if the system is asked "How did you determine the identity of ORGANISM-1? ", it answers by displaying each of the rules which were actually used, in the format shown above. This is something which the clinician can readily understand, and it provides a far more comprehensible explanation than would be possible if the program were to use a statistical approach to diagnosis.

It also means that the expert clinician can offer new "chunks" of knowledge, by expressing them in this same form. He can therefore help to make the program more competent, without having to know anything about computer programming.

There are several other interesting and important benefits gained from the approach we have chosen. These are explained in more detail in reference [1].

Progress report

General objectives and goals during the previous year

During the past year's work on the MYCIN system our goals have been (i) to increase the competence and broaden the scope of the system's therapeutic advice; (ii) to provide additional features to increase utility and performance of the system in the clinical setting; (iii) to develop further the system's collection of user-oriented features to make it easier for novices to use; (iv) to make it possible for an infectious diseases expert (who may know nothing about programming) to interact with and educate the program directly; (v) to develop new techniques to deal with the technical problems of managing a large and growing system; and (vi) to design and execute a formal evaluation study to measure the system's performance. We consider each of these in turn.

Competence and scope

One of the major accomplishments of the past year was the extension of MYCIN to cover diagnosis and therapy for meningitis infections. Over 100 new rules were added to provide this capability. This has proved to be an especially useful new domain to investigate because it has presented several new challenges. In particular, meningitis requires the ability to deal with a disease that is often diagnosed on clinical grounds before any specific microbiologic evidence is available. We have thus found it necessary to consider a larger range of clinical factors. This has resulted in a system which has a broader picture of the whole patient, and thus directly confronts one of the concerns about earlier versions of the system. The system has also become more robust, because it requires less hard microbiologic data, and is thus less sensitive to inaccurate laboratory reports. Like expert clinicians, it is now alert to the possible existence of anomalous data.

The broader range of expertise also means that the MYCIN can begin to play a much more effective role in the clinical setting. Another early concern was that a system with too narrow a range of capabilities demands a great deal of judgment before it is even used. Thus, if MYCIN could deal only with bacteremia, the user would need to decide that the patient indeed had a bacteremia before he could use the system. By giving MYCIN the ability to diagnose and treat a broader range of infections, we allow it to become useful at a much earlier stage in a patient's clinical course.

Other contributions to the system's competence came from the expansion of the knowledge base to include information about normal flora for a wide range of culture sites. MYCIN can now usually distinguish between normal and pathological flora, and can hence decide more precisely whether to treat.

We have also investigated the addition of some widely applicable routines for computing drug dose in renal failure. These have been developed by independent investigators, but are available to us and could prove to be extremely useful. Our system currently issues warnings simply to modify dosage in renal failure. Since the problem of determining renal status and the proper adjustment of drug dose is difficult, customized drug dosage recommendations will be an important addition to the power of the system.

There have also been significant improvements in the system's ability to handle organism genus and species. The problem requires that the system be able to deal with varying degrees of specificity; at times it can deduce both genus and species, and at others only the genus. Yet it must be able to prescribe correctly in all cases. A fundamental review of the problem has resulted in the addition of a number of new rules which handle the problem comprehensively and uniformly.

Additional clinical features

Several new features have been added to the system in anticipation of its use on the wards. MYCIN now keeps continuous statistics on the use of individual rules from its knowledge base. This will help us to monitor long term performance, to study interrelationships between rules, and perhaps detect inconsistencies or gaps in the knowledge base.

Also looking ahead, we have designed an "on-line" evaluation. At the end of each consultation, the system will ask a few questions about quality of performance, to get some feedback from the clinicians who are actually using it. This interchange will be very brief to avoid being a burden to the user, but will offer a very important form of instant criticism from our users.

User-oriented features

Several "human engineering" capabilities have been improved over the past year. For instance, the system's handling of questions asked by the physician has been made more powerful. This was achieved by improving our handling of English text, and by a comprehensive review of the kinds of questions that are asked. The system can now answer a broader range of questions, and, in particular, can explain why it did not take a specific action, as well as why other conclusions were reached. Capabilities like these are very important in allowing our local clinical experts to discover the program's rationale for its actions. They can then evaluate its line of reasoning, and suggest any necessary changes.

We are also engaged in a comprehensive effort to put all of the system's deductive actions into rules. Some important steps were previously performed by blocks of code, and hence could not easily be explained by the system. We have begun to reformulate the process in terms of rules. This will permit the system to be more specific about the source of its drug recommendations.

We have also added several new capabilities to provide more convenient use of the system in anticipation of its use on the wards. Among these are the ability for the user to type a comment about system performance at any time during the consultation. His comment is recorded in a special file which is periodically reviewed by our medical staff. This is in addition to the "on line" evaluation described above, and allows the user to offer any comment which he may feel is relevant.

We also have a parallel ability to report problems. The user can indicate that the system has "broken down" in some way, and is invited to describe the problem. His description is saved along with a copy of the program, so that our systems programmers can fix it later.

Linking the expert and the program

We have recently implemented a prototype version of a "bridge" between the clinical expert in infectious disease and the program, which will allow the expert to "teach" the program directly. Formerly, the expert's comments on the system's performance were given to a programmer, who then made the relevant changes to MYCIN. Now the expert can himself begin to discover the source of many problems, and can indicate the necessary rules. The dialogue is carried out in English, and requires no knowledge of programming.

Technical issues

Several changes in the structure of the program have made it easier to deal with the large and constantly changing knowledge base. In general, we are faced with the challenge of keeping the system's size within well specified limits, and have devoted some effort to insure that it remains sufficiently compact. We have, for instance, separated MYCIN's dictionary of English words from the rest of the system. This not only reduces the space requirement considerably, but has an additional benefit of making it easier to update the dictionary as the system grows.

There have always been extensive "self-documenting" capabilities in the system that is, MYCIN can supply instructions and helpful information if the user is confused at any point. We have recently improved the handling of this feature so that it is both faster and requires less space.

Formal evaluation study

A major undertaking this year has been the design and execution of a formal evaluation of the system's performance. The basic idea was to give the same clinical data to both MYCIN and a set of recognized experts in infectious disease therapy, to compare their judgments, and to ask the experts to evaluate MYCIN's performance. We began by designing a form that would allow us to separate the variables requiring analysis. We attempted to determine whether MYCIN (1) asks too many or too few questions, (2) correctly determines which infections require treatment, (3) correctly identifies the organisms that may be causing the relevant infections, and (4) adequately selects therapy to cover for the relevant organisms. The form was designed to be maximally informative, but very simple to complete. It interweaves a sample consultation with questions to the expert, and asks him to record his own opinions regarding the patient and appropriate therapy It was tested first in a pre-evaluation trial run, with five patients evaluated by three local Fellows in the Division of Infectious Disease at Stanford.

For the formal study, fifteen patients were selected according to strictly defined criteria. For each of these patients we prepared a 1-2 page clinical summary and made copies of relevant material from the patient's chart. This information was used to obtain therapeutic advice for each of them. Questions posed by MYCIN were answered solely on the basis of information collected from patient records at the time of the first positive blood culture, to simulate actual clinical use of the system. These consultations were integrated into the forms and sent along with the clinical data to ten experts in infectious disease therapy.

We had decided some time ago that the introduction of the system onto the wards for experimental use would be predicated on a successful outcome of this evaluation. Thus, while we had originally expected to begin use on the wards some time this year, the large amount of work involved in carrying out the evaluation has delayed us. We feel quite strongly that premature introduction of the program would be unwise, since it would almost surely lead to reduced acceptance by the clinical staff. Upon the return of the evaluation forms in mid- to late-March we shall have sufficient data to determine not only the current level of MYCIN's performance, but also the degree of agreement among the experts themselves. By sending five of the ten evaluation packets to experts in other parts of the country, we are also attempting to determine to what extent MYCIN reflects clinical judgments that may be peculiar to the Stanford environment.

In summary, our work in the past year has focussed on broadening the scope of clinical competence of the system, and on evaluating its performance. In anticipation of the use of MYCIN on the wards, we have added and strengthened many features, to insure that the program is maximally useful to the clinician who seeks advice.

Plans for the coming year

There are a number of major projects planned for the coming year.

There will, for instance, be extensive testing of the new meningitis rules, to insure both that their performance is satisfactory, and that there are no unforseen side effects on the rest of the system. We plan also to begin work on a knowledge base for pneumonia as the next step in increasing the program's scope. The introduction of the system onto the wards will give us valuable experience on a wide range of cases, and provide a basis for on-going monitoring and evaluation of performance.

We plan also to restructure part of the program's approach to requesting information from the user. Our current technique has developed a small number of technical problems, the most important of which involves the order in which questions are asked. By reorganizing some aspects of the control structure slightly, we will be able to solve all of the technical problems. From the user's point of view the system will continue to function as before, but at the start of a consultation it will ask a number of questions to get a global picture of the current case. This offers the additional, unanticipated, advantage of making interaction with the system seem more natural to the user who is used to presenting the consultant with a brief overview of the case at hand.

We have recently discovered an increasing need for the ability to use information about one infection to conclude things about another, as for instance when one infection has clinical implications about another. We plan in the coming year to implement this capability in a quite general fashion, so that we can deal with interrelationships of infections, cultures, organisms, and so on.

As the program becomes available on the ward, it will become more important to be able to tell the system about new information which may arrive several days after the initial consultation. Thus, as new test results arrive, or as the patient's condition changes, it should be possible to add the new information, and obtain updated recommendations. We plan to implement this too in the coming year.

Finally, since one of our fundamental tasks is the assembly of large amounts of knowledge of infectious disease diagnosis and therapy, we plan to develop further the prototype "bridge" which links the medical expert with MYCIN. Our current version lacks in particular many of the "human-engineering" aspects which are so extensively developed in the rest of the system. We foresee an important acceleration of this process of knowledge gathering when it becomes easy for an expert by himself to make significant changes to the knowledge base.

References

[1] Davis R, Buchanan B, Shortliffe E H, Production Rules as a Representation for a Knowledge-based Consultation System, A.I. Memo 266, Stanford University, November 1975. (submitted to Artificial Intelligence).

[2] Kagan B M, Fannin S L, Bardie F, Spotlight on Antimicrobial Agents - 1973, JAMA, 226, 3 (October 1973) pp 306-310.

[3] Neu H C, Howrey S P, Testing the Physician's Knowledge of Antibiotic use, NEJM, 293, 25, (18 Dec 75), pp 1291-5.

The MYCIN Project - List of Publications

[1] Scott A C, Clancey W, Davis R, Shortliffe E H, Explanation capabilities of knowledge based production systems (in preparation).

[2] Shortliffe E H, Davis R, Some considerations for the implementation of knowledge-based expert systems, SIGART Newsletter, 55:9-12, December 1975.

[3] Shortliffe E H, Computer-Based Medical Consultations: MYCIN, (adaptation of thesis), American Elsevier, New York, 1976

[4] Davis R, Buchanan B, Shortliffe E H, Production rules as a representation for a knowledge-based consultation system A.I. Memo 266, Stanford University, November 1975. (submitted to Artificial Intelligence).

[5] Davis R, King J J, An Overview of Production Systems Machine Representations of Knowledge, Proceedings of NATO ASI Conference, to appear, Spring 1976. (Also A.I. Memo 271, Stanford University, October 1975).

[6] Shortliffe E H, Judgmental knowledge as a basis for computer-assisted clinical decision making, Proceedings of the 1975 International Conference on Cybernetics and Society, pp 256-7, September 1975.

[7] Shortliffe E H, Axline S, Buchanan B G, Davis R, Cohen S, A computer-based approach to the promotion of rational clinical use of antimicrobials, International Symposium on Clinical Pharmacy and Clinical Pharmacology, Sept 1975, Boston, Mass. (invited paper)

[8] E H Shortliffe, R Davis, S G Axline, B G Buchanan, C C Green, S N Cohen, Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system Computers and Biomedical Research, 8:303-320 (August 1975).

[9] E H Shortliffe and B G Buchanan, A Model of Inexact Reasoning in Medicine, Mathematical Biosciences 23:351-379, 1975.

[10] Shortliffe E H, Rhame F S, Axline S G, Cohen S N, Buchanan B G, Davis R, Scott A C, Chavez-Pardo R, and van Melle W J, MYCIN: A computer program providing antimicrobial therapy recommendations (abstract only). Presented at the 28th Annual Meeting, Western Society For Clinical Research, Carmel, CA, 6 Feb 1975. Clin. Res. 23:107a (1975). Reproduced in Clinical Medicine, p. 34, August 1975.

[11] Shortliffe E H, MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection (abstract only);

Proceedings of the ACM National Congress (SIGBIO Session), p. 739, November 1974. Reproduced in Computing Reviews 16:331 (1975).

[12] E H Shortliffe, S G Axline, B G Buchanan, S N Cohen, Design considerations for a program to provide consultations in clinical therapeutics, Presented at San Diego Biomedical Symposium 1974 (February 6-8, 1974).

[13] E H Shortliffe, S G Axline, B G Buchanan, T C Merigan, S N Cohen, An artificial intelligence program to advise physicians regarding antimicrobial therapy, in Computers and Biomedical Research, 6:544-560 (1973).

[14] Shortliffe, E H, MYCIN: A rule-based computer program for advising physicians regarding Antimicrobial therapy selection, Thesis: Ph.D. in Medical Information Sciences, Stanford University, Stanford CA 409 pages, October 1974. (available from NTIS as document ADA001373)

Funding Status

MYCIN is funded by the Bureau of Health Services Research and Evaluation, and is currently completing the second year of a three year grant (S. Cohen & S. Axline, Principle Investigators). The budget for the current year (6/1/75 - 5/31/76) is $163,965; a total of $149,982 is requested for 6/1/76 - 5/31/77. Renewal for the coming year is currently under-going an in-house review. A new 3-year grant request for comparable levels of funding will be submitted in the fall of 1976.

Interaction with SUMEX-AIM Resource

During the past year, we have been contacted by a number of physicians who had read about MYCIN and were interested in trying out the system. Using TYMNET, these physicians in Boston, San Diego, Seattle, Washington D.C., and Atlanta were able to use the SUMEX GUEST account to run consultations on test cases.

MYCIN users are urged to send us comments about the system's performance. [A new "COMMENT" feature in the system allows comments to be entered at any time, without interrupting the consultation, nor even requiring the user to know how to use SNDMSG.] Recent comments from doctors associated with Rutgers Computers in Biomedicine served as very helpful guidelines for making the program's instructions and questions easier for a naive user to understand. Such comments also focus our attention on deficiencies in the program's medical knowledge as well as pointing out programming problems which may exist.

We have continued interaction via SNDMSG and terminal links with members of the DIALOG group, who recently wrote MYCIN-like rules for diagnosing and treating venereal diseases. We have implemented these rules and can modify or add to them as the doctors in Pittsburgh run more consultations to test the validity and completeness of this set of rules.

At a resent 3 day mini-conference, and at weekly meetings, members of the different SUMEX-AIM projects which make up the Heuristic Programming Project at Stanford discuss common problems faced by all the groups, and how each group handles these problems. These discussions have proved very helpful, especially to those projects which are currently in the design stage. Several of the projects have been able to benefit from the work that has been done in MYCIN on designing production rules and explanation capabilities.

Critique of Resource Services

Development of MYCIN has been greatly facilitated by the availability of the Interlisp language and its extensive interactive debugging capability and user-oriented features; in fact, it is doubtful that MYCIN could have developed to its current state without this large-scale interactive resource. However, in recent months its use has been severely limited by the poor response time during peak hours, which effectively prevents the use of MYCIN or Interlisp at such times. In this regard, we have found useful the SUMEX batch facility, which permits us to run some of our non-interactive tasks at times of low system usage.

We are fairly pleased with the availability of disk storage, although its availability may, in the foreseeable future, present some problems. Continuing development of the project makes substantial demands on disk space, since both experimental and publically accessible version of the system must be kept available, as well as a library of patient cases, system dictionaries, and documentation. The archival and retrieval mechanism currently available has proved to be very helpful, and we have made considerable use of it. This, along with careful management of the available space, has made it possible to avoid any problems at this time. As project development continues, however, we anticipate that disk space may become a scarce resource.

One of the outstanding aspects of the facilities at SUMEX continues to be the attitude and competence of its staff and systems people. They seem constantly willing to help with problems and consider suggestions, encouraging a sense of cooperation in the user community. Their on-going development of text editors and other features of the system contribute directly to its utility as a scientific resource.

IV.A.1.c PROTEIN STRUCTURE MODELING PROJECT

Protein Crystallography Project

Prof. J. Kraut and Dr. S. Freer (Chemistry, U. C. San Diego) and
Prof. E. Feigenbaum and Dr. R. Engelmore (Comp. Sci., Stanford)

(Grant NSF DCR74-23461, 2 years, $88,436 total)

I. Summary of Research Objectives

A. Collaboration, Biomedical Relevance and Technical Goals

The Protein Crystallography Project is a collaboration of two research groups, one at Stanford University, the other at the University of California at San Diego. The Stanford group consists of Edward Feigenbaum, Robert Engelmore, Penny Nii, and, during the current academic year, Carroll Johnson of Oak Ridge National Laboratory. The primary activities are to 1) identify critical tasks in protein structure elucidation which may benefit by the application of AI problem-solving techniques, and 2) design and implement programs to perform those tasks. The UCSD group consists of protein crystallographers: Joseph Kraut, Richard Alden and Stephan Freer. As protein crystallographers, their objective is to seek new techniques that will facilitate the elucidation of the tertiary structure of proteins.

The biomedical relevance of protein structure determination is well known. Solved protein structures have contributed substantially to our understanding of molecular biology, enzymology and immunology.

We have identified two principal areas where we believe the collaboration is of practical and theoretical interest to both protein crystallographers and computer scientists working in AI. The first is the problem of interpreting a 3-dimensional electron density map. The second is the problem of determining a plausible structure in the absence of phase information normally inferred from experimental isomorphous replacement data.

B. Specific Objectives

1. Interpretation of electron density maps

A major challenge in protein crystallography is the interpretation of the crystallographic electron density map. Our goal is to build a knowledge-based system which proposes plausible locations for substructural units consistent with the electron density map, the amino-acid sequence (if known), and physical, chemical and stereochemical constraints. The system attempts to integrate knowledge from three different areas: chemical topology, microstructure and macrostructure. The chemical topology knowledge base is task specific and contains the